

**TITLE**

**SPLIT GATE FLASH MEMORY DEVICE AND METHOD OF FABRICATING THE  
SAME**

**BACKGROUND OF THE INVENTION**

5      **Field of the Invention**

The present invention relates to a semiconductor device and method of fabricating the same. More particularly, it relates to a split gate flash memory device that can achieve high memory cell capacity.

10     **Description of the Related Art**

Complementary metal oxide semiconductor (CMOS) memory is generally categorized into two groups: random access memory (RAM) and read only memory (ROM). RAM is a volatile memory, wherein the stored data is erased when power is  
15     turned off. On the contrary, turning off power does not affect the stored data in a ROM.

In the past few years, market share of ROM has been continuously expanding, and the type attracting the most attention has been flash memory. The fact that a single  
20     memory cell is electrically programmable and multiple memory cell blocks are electrically erasable allows flexible and convenient application, superior to electrically programmable read only memory (EPROM), electrically erasable programmable read only memory (EEPROM) and programmable read  
25     only memory (PROM). Furthermore, fabricating flash memory is cost effective. Having the above advantages, flash memory has been widely applied in consumer electronic products, such as digital cameras, digital video cameras,

mobile phones, notebooks, personal stereos and personal digital assistants (PDA).

Since portability of these electrical consumer products is strongly prioritized by consumers, the size of the products must be minimal. As a result, capacity of flash memory must increase, and functions must be maximized while size thereof is continuously minimized. Having an increased amount of access data, capacity of memory cells has been enhanced from 4 to 256 MB, and even 1G byte will become the market trend in the near future.

Hence, there is a need for a flash memory device with high memory cell capacity.

#### SUMMARY OF THE INVENTION

Accordingly, an object of the invention is to provide a split gate flash memory device that can achieve high integration of memory cells thereof.

Another object of the invention is to provide a method of fabricating a split gate flash memory device, wherein the size of memory cells thereof can be reduced.

Thus, a split gate flash memory cell in accordance with the present invention comprises a substrate having a trench therein and a conductive layer disposed on the bottom of the trench. A pair of source regions are each disposed in the substrate adjacent to one sidewall of the trench and electrically connected by the conductive layer. A source isolation layer is disposed on the conductive layer. A pair of tunnel oxide layers are each disposed on one sidewall of the trench and on the source isolation layer. A U-shaped floating gate having a plurality of inside tips is disposed on the source isolation layer and contacts the tunnel oxide

layers thereby. An inter-gate dielectric layer is disposed on the U-shaped floating gate. A control gate is disposed on the inter-gate dielectric. A conductive stud is disposed on the control gate and a drain region disposed in the substrate adjacent to the trench.

Furthermore, the method of fabricating the split gate flash memory cell in accordance with the present invention comprises providing a substrate, forming a plurality of parallel long trenches along a first direction in the substrate, forming a conductive layer and a pair of source regions on the bottom of each long trench, wherein the source regions are respectively disposed in the substrate adjacent to two sidewalls of each long trench and electrically connected by the conductive layer therein, forming a source isolation layer on each conductive layer, forming a tunnel oxide on two sidewalls of each long trench, forming a U-shaped floating gate with a plurality of inside tips and a connecting oxide layer therein on each source isolation layer, forming an inter-gate dielectric layer on each U-shaped floating gate and the connecting oxide layer therein, forming a control gate on each inter-gate dielectric layer, forming a conductive stud on each control gate, forming a plurality of parallel shallow trench isolation (STI) regions along a second direction, defining a plurality of cell trenches, and forming a drain region in the substrate adjacent to each cell trench.

In the present invention, the trench-type split gate flash memory device disposed in cell trenches within a substrate can achieve higher integration of memory cell capacity than that in the Prior Art.

In addition, most of the fabricating processes in the invention are self-aligned and additional lithography processes and number of masks for the whole fabricating process can be reduced. The complexity of fabricating thereof is thus reduced and can be easily achieved.

Furthermore, most patterns of the masks for fabricating the stacked gate flash memory device are rectangular and can be easily fabricated. The costs of mask fabrication can be reduced and resolution limitations by the photolithography tools can be reduced.

A detailed description is given in the following embodiments with reference to the accompanying drawings.

#### **BRIEF DESCRIPTION OF THE DRAWINGS**

The present invention can be more fully understood by reading the subsequent detailed description and examples with references made to the accompanying drawings, wherein:

Fig. 1 is schematic top view of split gate flash memory devices of the invention;

Fig. 2a is a schematic top view of corresponding cross-sections in Fig. 3a and Fig. 4a of one embodiment of the invention;

Fig. 2b is a schematic top view of corresponding cross-sections in Fig. 3k and Fig. 4k of one embodiment of the invention;

Fig. 3a~3m are cross-sections of the fabricating process along the A-A' phantom line in Fig. 1; and

Fig. 4a~4m are cross-sections of the fabricating process along the B-B' phantom line in Fig. 1;

## DETAILED DESCRIPTION OF THE INVENTION

The present invention provides a split gate flash memory device that can meet demands for increased capacity of memory cells. In Fig.1, a top view of the split gate flash memory device in accordance with the present invention is shown. Structure of individual memory cells is disposed in the cell trenches (referring to trench 250') along the A-A' phantom line, between two shallow isolation trench (referring to STI) regions along the B-B' phantom line.

Fig. 3a-3m and Fig. 4a-4m respectively illustrate the cross-sections of a fabricating process along lines A-A' and lines B-B' according to an embodiment of the present invention. Moreover, Fig. 2a and 2b also illustrate corresponding top views of the fabricating process.

First, Fig. 3m illustrates a cross-section of the split gate flash memory cells in accordance with the present invention. A memory cell comprises a trench 250' disposed in the substrate 200, and a conductive layer 205 disposed on the bottom of the trench 250'. A pair of source regions S are each disposed in the substrate adjacent to sidewalls of the trench 250'. A source isolation layer 207 is disposed on the conductive layer 205. A pair of tunnel oxide layers 208 are each respectively disposed on one sidewall of the trench 250' and the source isolation layer 207. A U-shaped floating gate 209 having a plurality of inside tips is disposed on the source isolation layer 207 and contacts the tunnel oxide layers 208 thereby. An inter-gate dielectric layer 214 is disposed on the U-shaped floating gate 209. A control gate 215 is disposed on the inter-gate dielectric 214. A conductive stud 217 is disposed on the control gate

215, and a drain region D is disposed in the substrate 200 adjacent to the trench 250'.

In Fig. 3a and 4a, a semiconductor substrate 200, for example a P-type silicon substrate, is provided. Next, a pad oxide layer 201 and a mask layer 202 are sequentially formed on the substrate 200 and then sequentially defined by subsequent lithography and etching, forming a plurality of long trenches 250 with a depth about 30000Å to 70000Å in the substrate 200. The long trenches 250 are parallel to each other along a first direction. This top view is shown in Fig. 2a and the depth of a long trench 250 affects the channel length of each split gate flash memory cell.

Next, in Fig. 3b and 4b, a conformal bottom insulating layer 203 is deposited on mask layer 203 and in the long trench 250. The material of the insulating layer 203, for example, can be silicon dioxide. Then materials of a protecting layer 204 are deposited on the bottom insulating layer 230 filling the long trench 250. Material of the protecting layer 204, for example, can be photoresist. Materials of the protecting layer 204 are then etched back and recessed to a depth H to the bottom of the long trench 250. Then the bottom insulating layer 203 exposed by the residue protecting layer 203 is removed. Thus, a bottom insulating layer 203 having the same height H as the protecting layer 204 is left in the long trench 250.

In Fig. 3c and 4c, the residue protecting layer 204 in the long trench 250 is then removed by a proper solvent (not shown). Next, conductive materials of a conductive layer 205, for example N-type dopant doped polysilicon, are filled in the long trench 250 and on the mask layer 202. Then the conductive materials above the mask layer 202 are etched

back and the conductive materials in the long trench 250 are recessed to a depth  $H'$  to the bottom of the long trench 250. Thus, a conductive layer 205 with a thickness  $H'$  is left in the long trench 250. Moreover, a depth difference  $\Delta H$  between the bottom insulating layer 203 and the conductive layer 205 is about  $100\text{\AA}$  to  $200\text{\AA}$ .

In Fig. 3d and 4d, a source material layer 206 is formed in the long trench 250 by sequential deposition, etching and recession of doped materials. The doped material of the source material layer 206 can be, for example, N-type dopant doped silicon dioxide and preferably phosphorous (P) doped or arsenic (As) doped silicon dioxide formed by LPCVD. Thickness of the source material layer 206 is between  $500\text{\AA}$  and  $1000\text{\AA}$ .

Next, a thermal annealing process, for example a furnace annealing process, is performed on drive out dopants (such as arsenic) from the source material layer 206 into the substrate 200 adjacent to the long trench 250. Thus, a pair of source regions S are respectively formed in the substrate 200 and electrically connected by the conductive layer 205 therebetween.

In Fig. 3e and Fig. 4e, the source material layer 206 in the long trench 250 is then removed. Next, a source isolation layer 207 with a thickness between  $350\text{\AA}$  and  $850\text{\AA}$  is formed on the conductive layer 205, and limitation of an electrical connection between the source regions S is fully achieved by the conductive layer 205 therebelow. The source isolation layer 207 is formed by sequential deposition, etch-back, and recession of insulating materials such as silicon dioxide, for example. Next, a threshold voltage

implantation Vt is performed on sidewalls of the long trench 250 to adjust the threshold voltage of each memory cell.

In Fig. 3f and Fig. 4f, a tunnel oxide layer 208 is formed on the two sidewalls of the long trench 250. The tunnel oxide layer 208 can be, for example, silicon dioxide formed by thermal oxidation. Next, a conformable floating gate layer 209 is formed on the mask layer 202 and in the long trench 250. Materials of the floating gate layer 209 can be, for example, N-type dopant doped polysilicon formed by LPCVD. Next, a connecting oxide layer 210 with a thickness between 700Å and 1500Å is formed in the long trench 250. The connecting oxide layer 210 is formed by sequential deposition, etch-back, and recession of materials such as silicon dioxide. The connecting oxide layer 210 therein can lower the coupling ratio between the floating gate and the control gate of the flash memory cells of the present invention.

In Fig. 3g and 4g, portions of the floating gate layer 209 exposed by the connecting oxide layer 210 are removed, leaving a U-shaped floating gate layer 209 in the long trench 250 as a floating gate. Then a conformable floating gate oxide spacer layer 211 is formed on the mask layer 202 and in the long trench 250 and etched, leaving floating gate oxide spacers 211 covering part of the tunnel oxide 208 on sidewalls of the long trench 250. The conformable floating gate oxide spacer layer 211 is formed by LPCVD, for example. In addition, materials of the floating oxide spacer layer 211 and the connecting oxide layer 210 are the same, such as silicon dioxide here. Thus, during etching of the floating gate oxide spacer layer 211, the connecting oxide layer 210 is lightly etched and the thickness thereof is also



decreased by the described etching. Then a floating gate nitride spacer 212 is formed on each floating gate oxide space 211 by sequentially depositing a conformable floating gate nitride spacer and etching thereon. The conformable floating gate nitride spacer layer 212 is formed by LPCVD, for example.

In Fig. 3h and Fig. 4h, the connecting oxide layer 210 can be etched, for example, by wet etching to further lower a thickness about 200Å to 500Å thereof by exposure of the floating gate nitride spacers 212 and insides of the U-shaped floating gate 209. Next, the insides of the U-shaped floating gate 209 are isotropically etched and a plurality of tips is formed thereon. Thus, in the long trench 250 a U-shaped floating gate 209 with a plurality of inside tips and a connecting oxide layer 210 therein is formed on the source isolation layer 207 therein.

In Fig. 3i and Fig. 4i, the floating gate nitride spacers 212 can be removed, for example, by wet etching. Then, another protecting layer 213 with a thickness between 500Å and 1000Å is formed in the long trench 250. Materials of the protecting layer 213 can be, for example, photoresist (PR). Then wet etching removes the floating gate oxide spacers 211 and portions of the tunnel oxide layers 208 from the sidewall of the long trench 250 using the protecting layer 213 as a masking layer.

In Fig. 3j and Fig. 4j, the protecting layer 213 is then removed using suitable solvent. Next, a conformable inter-gate dielectric layer 214 formed by LPCVD, for example, can be deposited on the mask layer 202 and in the long trench 250. The material of the inter-gate dielectric layer 214 can be, for example, silicon dioxide. Then a

control gate layer 215 is formed in the long trench 250 having a height slightly lower than that of the surface of the substrate 200 as a control gate. The control gate layer 215 is formed by sequential deposition, etch-back, and recession of materials like N-type dopant doped polysilicon formed by LPCVD, for example.

In Fig. 3k and Fig. 4k, a conformable layer of control gate spacer layer 216 is deposited and then etched by subsequent etching. Thus, a control gate oxide spacer 216 is formed on each sidewall of the long trench 250. The material of the control gate oxide spacer 216 can be, for example, silicon dioxide. Next, materials of a conductive stud 217 are deposited on the mask layer 202 filling the long trench 250 followed by etching. Thus, a conductive stud 217 is formed in the long trench 250 as a connection of the control gate (referring to the control gate layer 215) therebelow and a wordline formed by subsequent processes (not shown).

Next, a plurality of parallel long isolation trenches are formed in the substrate 200 (shown in Fig. 4K) along a second direction, perpendicular to the first direction of the long trenches 250 by sequential lithography and etching. The described etching process stops at the source isolation layer 207 of the long trench 250 therein and a plurality of trenches 250' comprising the cell structures is thus defined in the substrate 200. This top view is shown in Fig. 2b. Fig. 3k shows a cross-section along the A-A' line where the trenches comprising cell structures are located and Fig. 4k shows a cross-section along the B-B' where follow-up shallow trench isolation (STI) regions are located.

In Fig. 3l and Fig. 4l, materials of an insulating layer 218 are filled in the described long isolation trenches. The insulating material can be, for example, silicon dioxide formed by high density plasma chemical vapor deposition (HDP CVD).

In Fig. 3m and Fig. 4m, a planarization process is then performed on level the wafer using the mask layer 202 as a polishing stop layer, leaving an insulating layer 218 therein and the shallow trench isolation (STI) region is thus formed. Next, the mask layer 202 is removed by, for example, wet etching, and the pad oxide layer 201 therebelow is exposed. Then a drain implantation (not shown) is performed on implant N-type impurities such as phosphorous (P) or arsenic (As) ions into the substrate 200. Then a thermal annealing process (not shown), for example a rapid thermal annealing (RTP) process, is performed and drain regions D are thus respectively formed in the substrate 200 adjacent to each trench 250'. Then the pad oxide layer 201 is removed and a second insulating layer 219 is formed on each drain region D. The second insulating layer 219 is formed by sequential deposition and planarization of materials of the second insulating layer 219. After the planarization, the wafer surface is leveled and the conductive stud 217 is thus exposed.

In Fig. 1, a top view illustrating possible follow-up wordline BL and bitline WL is shown. In Fig. 1, the conductive studs 217 within the cell trenches (referring to trench 250') are further connected by the interconnecting hypothesis wordlines WL and hypothesis bitlines BL in phantom structures. In addition, the bitlines BL can also connect the drain regions D therebelow by a proper contact

window (not shown) and the split gate flash memory devices are thus formed.

Compared with the flash memory cell of the Prior Art, the present invention has the following advantages.

5 First, cells of the flash memory device in accordance with the invention are trench-type split gate flash memory device, disposed in cell trenches within a substrate rather than those normally disposed on the surface of a substrate in the Prior Art. Memory cell design of the invention can  
10 thus achieve higher integration capacity than that in the Prior Art.

In addition, most of the fabricating processes in the invention are self-aligned. Thus, additional lithography processes and the number of masks for the whole fabricating  
15 process can be reduced. The complexity of fabricating the split gate flash memory device of the present invention is reduced and can be easily achieved.

Second, cells of the split gate flash memory device of the present invention are formed into the substrate. Thus,  
20 the size of each flash memory cell is minimized and the integration of the memory cell increased, allowing capacity of a flash memory device to be increased and the current within a cell also increased by enlarging the depth of the cell trench. Furthermore, most patterns of the masks for  
25 fabricating the stacked gate flash memory device are rectangular and can be easily fabricated. The costs of mask fabrication can be reduced as can resolution limitations by photolithography tools.

While the invention has been described by way of  
30 example and in terms of the preferred embodiments, it is to be understood that the invention is not limited to the

disclosed embodiments. To the contrary, it is intended to cover various modifications and similar arrangements (as would be apparent to those skilled in the art). Therefore, the scope of the appended claims should be accorded the  
5 broadest interpretation so as to encompass all such modifications and similar arrangements.